DIGITAL COMPUTER LABORATORY
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

Report No. 126

DECISION STRUCTURES FOR RECOGNITION

by

Albert H. Amon

October 1, 1962

This work was supported in part by Atomic
Energy Commission Contract AT(11-1) - 1018

TABLE OF CONTENTS

I. INTRODUCTION

The present report describes concepts underlying complex decision-making systems and how these systems handle certain information processing problems now easy for human beings but difficult for computers. The dominant issue discussed is the recognition of qualities of sensory input. It is likely that other inductive functions of the brain demand a similar analysis. In particular the concept of a link-forming recognition structure is introduced. Further the discussion of error detection, correction and avoidance suggests new decision organizations of enhanced reliability.

The problems of coordinating vast numbers of decisions made simultaneously have not been adequately faced either by psychologists or by logical designers. As a psychologist the writer claims no special competence in computer design. However, understanding these matters may allow the construction of computers of unexpected kinds.[1] It will be obvious that the inspiration for much of this presentation comes from psychology, here presented in a contrived translation into another language. No originality for the concepts presented other than the labor of assembly is claimed. This is an interim report that shows disturbing gaps; in particular the problems of development and discontinuous change in complex decision-making systems present challenges unanswered here.

[1] The work on the Illinois Pattern Recognition Computer is a promising beginning. Refinement of the present work may be of use in the later stages of control system design for this machine, and in the formulation of reliable pattern recognition programs.

II. SYSTEMS AND FUNCTIONS

A complex decision-making system[2] (S) may be characterized by its observable external functioning only, or in addition by its real or hypothetical internal processing. This report will speculate about possible models for internal processing available to complex decision-making systems. A functional description of a system, however, is provided by way of introduction.

A system is an assemblage of interdependent processors embedded in an environment to perform tasks.

It is supposed that S receives information from its environment, the input function of S. Input information may either be imposed in a fixed format upon the S by the environment, or the S may be allowed some selectivity in the rate it takes up input or in the order and kind of the input information itself.

The responses of S may either be outputs to the environment or internally induced changes. Input selection, if any, and response selection are to serve one or both of two major responsibilities of the S: short range reaction to environmental input appropriate to the tactical task; and longer range adaptation, producing changes in the S itself that make its future reactions more effective. Tasks may be externally imposed or set by the S to itself; in either case when they become firmly enough established to direct behavior of S they become purposes.

The reactive function of particular interest here is the recognition of environmental objects. Here reactions may be inconclusive trials through which additional environmental information is sought, and conclusions: recognitions called for by the task.

Conclusions may be evaluated by the environment, or by an internal evaluative process. When evaluations have a metric which allows comparison among concluded tasks (i.e. recognitions) they assign values. If this metric extends to evaluation of the resources expended by the S in obtaining

[2] Underlining is used here not for emphasis but to mark the first occurrence of terms that take on special meanings in this report.

particular values, then one can speak of the cost of obtaining them. Tasks for which it is possible to obtain certain values only at an increase in the cost of obtaining others conflict with the others.

Values may also be assignable to the achievement of adaptation, as may costs. Particularly difficult conflicts to resolve are those which occur between reactive and adaptive values, the measures of which may be incommensurate.

Recognition tasks may differ: in the degree to which S is to recognize anything present, identification; or in the degree to which S it to recognize only certain searched for things or events, detection; in the fineness of the necessary discriminations, critical resolution; in the cost of error or required accuracy; in the rate at which recognitions are to be produced, urgency; and in the degree to which this rate is determined by the environment or chosen by the S. S's which can adopt different reactive modes fitted to these or other task differences are flexible. Flexibility is not necessarily the same thing as adaptability; there may be conflict between the two. An example of this conflict would be adaptation cost which will occur when achieving certain strategic values prevents others from being realized.

For recognition tasks objects to be responded to (named) may be: discrete or intergrading; from a finite or infinite set; from a variable or fixed set; associated with spatially contiguous or dispersed aspects of the environment; and be themselves variable or not. The correspondence between object and name may or may not be a one-to-one mapping. The relationship may be such that certain relationships obtaining between the objects hold also between the associated names, in which case the naming is more or less regular. Naming may be arbitrary and allow no predictability of name for a novel object from a similar, previously named one. Naming in this case is arbitrary. Objects may or may not be characterizable by separable attributes (in ways that potentially facilitate recognition.) In some cases objects to be identified will increase in span, either in the amount of the environment involved or in the time span of relevance, as S itself develops. Context information, either from the domain or from previously stored information, may be utilized by the S to achieve recognition. Changes in the environment, its objects, or the names or other responses called for, may be predictable or unpredictable.

S may or may not have access to the information necessary for prediction. Further definitions of and distinctions between aspects of task and environment will be made as they contribute to comparisons between various S's.

III. PROCESSES AND PROCEDURES

A recognition task calls upon S to produce a conclusion. Trial responses may intervene which ask of the environment supplemental information--either additional samples from the domain, or guidance from another S, or both. Such elicitation of environmental feedback may be continuous or discrete. If the latter, it may be possible to divide performance of the task into problems, each problem representing a unit of processing which may be temporarily closed, pending further input from the environment. If such division is impossible, there is only one problem per task. A problem is reopened after receipt of appropriate information. Points of closing and reopening are transition points, and for a particular problem, make up transition pairs. If a task has only one problem, this may be reopened after evaluation indicates the response given was in error.

Should S find itself occupied with other tasks, the reopening of a problem may be delayed. If intervening activity does not make reopening of a problem any more costly, the closure is stable relative to that intervening activity. Stability of a closure may depend on the point chosen for reopening, and in this case is a property of the transition pair rather than of point of closure alone. It will generally be to the advantage of the S if closures are so chosen as to be stable when possible. If unstable closure is for some reason desirable, S may utilize caretaker processes, or residua, to minimize reopening costs.

Residua serve to maintain the readiness of S to reopen a problem. They may have additional monitoring functions such as maintaining a selective sensitivity to input information, cues, which could significantly affect the chances of success for the reopened problem. A residuum may monitor competing current tasks (committed processes) or even competing residua for its own success-promising cues, and the failure-promising cues of the competing processes.

The processing relevant to any problem is its procedure. Processes involved in a procedure may be variously grouped into subprocedures. If procedures of several problems make common use of certain subprocedures, they are shared. The sharing of common subprocedures is one way in which an S can realize values simultaneously which might require duplicated or serial effort in a less effective S.

S may simultaneously perform several subprocedures which may, but need not, be parts of the same procedure. This is parallelism. A subprocedure where no incremental progress is possible without corresponding incremental progress in other subprocedures is continuously dependent on these other processes. Subprocedures continuously interdependent on each other are essentially parallel. Whole procedures, if essentially parallel, carry out inseparable problems for which there is no distinction between sharing and parallel processing. Subprocedures, not essentially parallel, may be effectively carried out in parallel by S. S's differ in the number and kind of subprocedures that can be processed in parallel.

Parallel processing may present difficulties of coordination, the scheduling of subprocedures so that economy of simultaneous operation is not lost in confusion or error. Coordination difficulties may sometimes be reduced: by standardization and simplification of the operations involved; by distribution of control to separated units each with independent access to information necessary for appropriate decision; or conversely, by centralization of control when this is more effective. Better coordination sometimes results from a combination of these methods, for instance by appropriate selection of which decisions should be centralized, which distributed and to what extent, etc. Commensurable value measures for subprocedures may also aid coordination.

It is sometimes of value to perform in series subprocedures previously performed in parallel, either by completion of one before beginning another, or by alternation between the two. Such alternation may be of value when there are environmental or internal conditions that make the performance of one momentarily more efficient than the other, and when these conditions change over intervals too brief to allow selection of a particular most efficient order of performance. Alternation may also be

suggested when the relative difficulty to be encountered in performance of a subprocedures is not predictable. Here information resulting from alternation may facilitate planning. Alternation is a fine grained example of the use of closure and reopening, the effectiveness of which will often depend on the existence and detection of stable closure points or transition pairs in the subprocedures alternated. Development of further considerations of stability and residua formation will be deferred until later.

IV. REPRESENTATION OF STRUCTURE

A. Nodes and Channels

The structure of S at a given time is its internal constitution: its available facilities and the ways in which these can be organized. Structure includes immediate reactive modifications produced in A, in service of current purposes, but does not include longer range adaptive changes. The structure of an S may be unknown to the observer, but at any one time it has only one structure.

The functioning of S may be studied by the use of descriptive representations, of which there may be many seemingly consistent with what is known of S's structure.

For comfort in visualization, representations discussed here will employ facilities distributed in space. The unit of organization will be the node, regarded as having a complex of operations which process information and control informational access to, and exit from, the node. Information is seen as conveyed between nodes either by transfer along discrete channels or by diffuse transmission to all nodes sensitized to its reception. Because it is convenient to conceive of communication between nodes as confined to discrete messages, processes essentially parallel to each other will be assigned to the same node, as their coordination requires continuous intercommunication. Representation in terms of spatially localized nodes, intermittently communicating with each other, is generally applicable to S's. However, an S, extremely tightly coordinated, with every process essentially parallel to every other, might be describable as only one node,

an analysis of limited promise. The problem is one of degree: by ignoring weak interrelationships it may be possible to describe predictable performance of an S for interesting spans of time, and with adequate accuracy.

The choice of which processes are to be grouped into a node and which are to be separated into different nodes, is constrained only by the above convention concerning essential parallism. Internally complex nodes may be describable in terms of subnodes and their connecting channels, as complicated as the entire representation of other S's. At times it may be of value to consider S's as nodes of more comprehensive S's. Later discussion will argue that some kinds of grouping are more useful than others, and more likely to correspond to what naturally occurs in the structure of the S's studied.

Messages may involve much or little information. Substantive messages carry information about environmental objects or recollections of such objects from memory. Control messages carry information, usually less complex, which may affect operations at various nodes. This distinction will not always be meaningful, as the contents of a substantive message may be progressively disassembled into abstract attributes of an object, and these fragments transformed into control messages directing the processing of subsequent substantive messages. As a message is transfered from node to node it may have associated with it tags, or control messages, prepared in one node for use by another.

B. Operations: Tests and Transformations

Operations at a node may be roughly distinguished as either tests or transformations. Tests select the information they operate upon in terms of its origin or content, and control the destination to which this information is to be conveyed. Tests control the path followed by information within S. In this path control, then, tests are the basic decision-making elements of the S. A pure test directs, but does not alter, the information upon which it operates. A transformation on the other hand alters, but does not direct, the information on which it operates. The distinction is ultimately an arbitrary one, as path choice itself contains information, and what is done by a transformation in one representation of a S might be accomplished by a test in another.

For many purposes the distinction between input selecting and output selecting tests is an unnecessary one. Thus, it will often be possible to replace the input selection function of one test by the output selection functions of preceeding tests. Cases where the distinction is important will be met. But for present purposes the term "test" will mean an output selective test, unless otherwise specified.

Tests range in resolution from unequivocal selection of one alternative, through sequence selection enlightened by information concerning the probability distribution over outcomes, to arbitrary sequence selection.

Tests may differ also in the amount of time necessary for their completion. Sometimes, for progressive tests, the resolution of the outcome may be optional depending on transfer thresholds, on the time allowed for completion and on the amount of information utilized. In such cases the resolution required may differ from task to task and be controlled by information conveyed from other nodes. Changes in transfer threshold (in particular, differential threshold changes) can make some outcomes easier to test while making others more difficult, and therefore create a bias toward particular outcomes. Again, the order in which the outcomes are to be tried may be influenced from other nodes.

Transformations may variously affect messages operated upon. Information may be reproducted and the message so duplicated sent simultaneously over several paths for parallel processing. A transformation may affect the message irreversibly, in which case reproduction preserves the original message. Information may be substituted for message information in a regular or arbitrary way (the same terms as used earlier for the naming transformation). The substitution may be complete or partial. A transformation may be fabricative in that it adds to a message information for some reason missing. An example of fabrication is the filling in of the retinal blind spot in such a way that a human S normally remains unaware of the addition.

Transformations may disassemble a message into submessages, each containing only part of the original information; others reassemble such a divided message after various vicissitudes of processing.

Transformation processes are not limited to change of substantive messages. Change of operations at a node may occur also, as a result of transformations triggered by control messages. Control modifications possible include: alterations of transfer thresholds at nodal tests; alteration of the arrangement of sequencing nodes; and opening and closing of various channels, even to complete bypassing of certain tests and transformations.

Of particular interest for contextual synthesis and for error correction is the observation that certain transformations may be incompatible, if environmental laws make it impossible for two particular messages to be such that one could be transformed one way, the other another way. An example of such incompatible transformations would be making one of two objects larger, and the other unaltered or smaller, when they are known to be at the same distance from the observing S and their relative sizes known and fixed.

V. PROCEDURE ORGANIZATION IN RECOGNITION STRUCTURES

In this section some alternative models of recognition structures will be described briefly. Additional detail will be supplied in later sections, where there will be relaxation of simplifying assumptions adopted here. Considerations of relative effectiveness of these structures and of possible modification strategies will also be postponed until then.

A. Parallel Matching

For the moment it is assumed that each of the objects to be recognized belongs to only one of a number of discrete types. An input message regarding the object is a sample. The recognition procedure consists then of pairing each sample with the name of a particular type.

A very simple recognition procedure involves simply comparing the sample with stored information characterizing each type. This procedure might be quite efficient if there were facilities for simultaneously matching each sample against each possible type. Although some parallelism in matching may be possible, provision of such facilities will generally be prohibitive if the number of types to be recognized is large. More

efficient recognition procedures will select more likely types before matching is attempted. Each type may correspond to objects and samples widely differing in characteristics, many irrelevant to the discrimination. Attempted matching of all possible type variations would greatly tax matching facilities. It may be more efficient in these situations to transform the sample into as invariant a form as possible before attempting a match. As many types may require the same transformations, and the selection of possible transformations made by identical tests, it may be inefficient to assign independent facilities for transformations for each type. The limitations of a purely parallel matching structure have been discussed here as if matching were integral to the recognition. Recognition structures, however, may not use matching at all.

B. Progressive Tracing Structures

Structures of greater organizational depth will now be considered. The organizational relationship established between nodes by interconnecting channels in the connectivity of the structure. Transient organizations which may occur as certain channels are activated or closed off are linkages. First, divergent detection trees will be discussed. "Tree" here is used as in the theory of abstract graphs, to designate structures with no more than one path connecting any two nodes.[3]

In progressive tracing structures samples are introduced only at the root node, the node of rank zero. Procedures continue at successive nodes, the rank of each of which is the rank of the immediately preceeding node, plus one. The termini of the tree are the nodes of highest rank on their respective branches.

The simplest progressive tracing models use undistributed sample structures. In these the sample is transfered from the critical node at which the sample is being tested to only one of the nodes of next higher order. Terminal nodes of the tree correspond to types to be distinguished. Any information in the sample not required for the tests used at preterminal nodes is supplementary. Termini which have associated with them information about the types, other than that used in the preterminal tests, have

[3] As used in this report information flow in a tree need not be exclusively from root to termini. For example, in later discussing echo techniques and error retraces, signal flow from terminal toward the root is explicitly introduced.

type memories. *Terminal checking* trees have tests assigned to terminal nodes which compare the type memories with the supplementary information of the sample.

The next simplest progressive tracing models are *distributed sample* structures. In these the sample is reproduced at all or some nodes so that the sample may be distributed simultaneously to more than one node of higher rank. It is still assumed that only one of the terminal types corresponds to any object to be recognized. Disbributed sample structures have tests assigned to nodes to reject the sample, sooner or later, along all branches other than the one leading to the correct terminus. If several of the reproduced samples reach termini, the terminal tests should eliminate all but one. If the sample is sent up all branches and rejection occurs only at terminal nodes, this model reduces to the parallel matching model. Distributed sample structures can also approach the undistributed sample model, if only few nodes transfer the sample to more than one node, and for all nodes, except on the correct branch, rejection is quick. As a model approaches the undistributed sample extreme, it is said to be more *selective*.

Progressive tracing models need not be limited to models completing one recognition procedure before beginning another. In undistributed sample structures only one of the processing nodes will be active at any time. By giving each sample an environmental *location* tag, and by providing buffering facilities to prevent interruption of active nodal processes by incoming samples, *multiple sample* structures are possible. The location tag provides for the reassembly of the objects recognized in their environmental order. Multiple sample structures can either maintain a constant load, by admitting a new sample to the tree for each sample achieving recognition, or take up samples in batches, perhaps from an environmentally related object set. This technique can facilitate recognition by use of contextual information, should some of the types first recognized in the multiple sample structure supply constraining information fed back to nodes still processing other traces. (For some applications, see Sec. VI D)

Progressive tracing structures may have certain nodes which are always sequencing nodes to provide for an orderly testing of alternatives. Sequencing nodes may be regarded as nodes for which an appropriate

test has yet to be found. A node may revert to a sequencing operation, if tests assigned it are powerless because a particular sample lacks the attributes tested. Nodes always sequencing in function generally are better placed near the termini of a structure to allow search among possible alternatives to be more readily accomplished. Later discussion of echo-processing, retracing and error detection will illustrate reasons why additions to a structure are more readily made at high ranking nodes.

Where a sequencing node, or one that must frequently serve as such due to sample insufficiencies, must be placed in a low ranking position, it may be of value to provide for *relative*, preterminal *closure*. This closure may be accomplished by the technique of *admittance testing*: tests to determine whether or not the correct branch has been entered. If such testing is distributed over several successive nodes, relative closure is said to be *completed* at the highest ranking node of the admittance testing series. Of course admittance tests could be performed after tests at the sequencing node, but this might not allow efficient simultaneous testing in parallel distributed S models. In any model, testing at later nodes may be facilitated by lower-ranking admittance testing, as in each branch there may afterwards be less to test than previously at the original ambiguous node. Also, admittance testing may be of value as a continuous check in structures, no one of which may be conspicuously uncertain. (See Sec. 6 D, error location.)

C. Fore-testing, Echo-techniques, and After-Processing

Structure connectivities discussed have been simple trees with undirectional passage of information from root to termini. Sample disassembly can occur at the root or subsequent node, and partial samples sent ahead in the tree to allow partial testing of these *forerunners*. On the basis of the outcomes of forerunner testing, the processing of the entire sample might be facilitated (i.e. by the deletion of tests upstream.) Advantages of forerunner tests include: trial *anticipatory* tests by matching against more likely types, freeing the probably irrelevant parts of the structure for multiple sample processing, and enabling a general Gestalt to form, in terms of which the progress of subsequent more detailed testing might be judged and possible errors forestalled. If forerunner processing produces a simplified structure by making some tests unnecessary through the closing of channels, it is *fore blocking*.

In some situations partial matching amy be more desirable than complete matching, either in speed or in permitting more completely parallel processing. If the undirectional constraint on the flow of information in the tree is relaxed echo techniques, whereby the outcome of a partial matching attempt is conveyed back to the echo inducing node (often but not necessarily the root), can be imagined by making use of backwardly propagated, or reflected echo messages. Back blocking is the closing of all channels to higher ranking nodes other than those over which the echo was received. Because of the partial nature of echo matching, a number of termini will in general reflect echos. In such cases back blocking produces a reduced tree in which necessary critical testing can be accomplished with fewer decisions. Echos may be reflected from termini only, or from intermediate nodes where relative closure is possible (as by admittance testing). Echo structures can also differ in the way sample information is conveyed to potentially reflecting nodes. This information need be no more complete than required by the partial matching processes which determine what will be reflected. If this information is transfered through the tree by the normal sample message channels, it becomes a special kind of forerunner processing, where no decisions are made until reflecting nodes are reached. If intermediate message forerunner testing is used to limit the nodes from which reflections will be possible, then echo and forerunner processing are combined. However, it is not necessary that the sample information be so transfered to the reflecting nodes. Direct input to reflecting nodes through special, simple input channels, or even transmission may be employed.

The echo signal reflected back down[4] the tree may be nothing more than a pulse necessary for back blocking. Echo magnitude can be used to encode changes in the order of sequencing nodes or to change test thresholds. In structures other than trees, the echo might acquire information as to path taken as it passed each node on the way back. More complicated possibilities include the progressive echo technique. Let the echo contain information (e.g. from terminal type memory) other than that transmitted

[4] That is, toward the root of the tree.

to it for partial matching. Matching this transformed echo to the sample at the echo inducing node could confirm the echo, discriminate the more promising from among candidate returned echos, or even start a new echo cycle involving more comprehensive, or in any case different, partial matchings.

Both forerunner and echo techniques involve partial testing on partial samples, disassembled and distributed in the structure. They produce reduced trees which are intended to allow a simplication of subsequent critical testing. Auxiliary testing processes may lag behind rather than lead the critical processes. An example is after-testing, in which the testing processes of a node continue to process the same sample after it has already been passed on to subsequent nodes. In this case there is a difference between transfer thresholds of a test and more rigorous closing thresholds. After-testing will be considered later (in Section VI) in connection with error avoiding techniques.

Tree simplification, either by fore or by back blocking, has been described as though limited to omission of nodes and closing of channels. However, internal nodal test and transformation structure can also be modified to simplify the critical procedure. Thus that part of a nodal operation that serves only to distinguish between alternative blocked channels can be eliminated, as can any part which distinguished only between blocked and unblocked channels. Operations distinguishing between still unblocked channels alone need be kept. However, a limit may soon be reached in the effectiveness of providing many alternative sets of operations for each node, one set for any combination of possibly unblocked channels. It should be noted that improvements in efficiency, possible by these structure-simplifying techniques, may be accompanied by an increased vulnerability of the S to error, should the correct type belong to one of the eliminated parts of the structure.

D. Other Structures

Divergent trees so far considered produce multiple outcomes only as different objects are recognized. Different in purpose are attribute-isolating trees where attributes of one object rather than types are distinguished. In such a structure nodal tests control the disassembly

of a sample into sub-samples, passed on to different branches. Ultimately, each attribute is identified at a terminus. A subsequent convergent structure can reassemble the idealized sample into a form appropriate for a later recognition structure. The two structures together perform a complex transformation of the sample.

Structure connectives discussed have been simple trees. Retaining this limitation on connectivity for channels conveying substantive messages, but relaxing it for control connectivity, more interesting organizations become possible. Thus information resulting from operations in other parts of the structure, (i.e. not necessarily on the same branch), can be utilized in influencing operations at a node. Procedures having achieved, or nearly achieved, recognition could in this way feed back contextual information to supply helpful constraints for more slowly progressing procedures. Control information need not arise within the decoding tree. Control information from outside the decoding tree may modify the tree to facilitate special detection problems [for example, by affecting sensitivities]. Particular uses of this environmental control will be examined later, after some possible mechanisms have been suggested here.

The possibility of utilizing control information derived from parts of the S outside the immediate detection structure has been mentioned. One method of some promise stems from the observation that whether types are to be regarded as different, or alternatively, as essentially the same, may depend on the current, dominant purpose of the S. Thus classes of types may be purpose-equivalent. The echo techniques of the back blocking may simplify the tree detection by bypassing all tests serving only to distinguish between purpose-equivalent types.

VI. USES OF REACTIVE MODIFICATION IN DETECTION

The utilization of reactive modification in the service of detective purposes of the S will be considered here. For the sake of exposition relatively transient changes will be distinguished from adaptive changes which though not necessarily realized immediately yield an enduring, improved effectiveness of the S. It is assumed

for this report that the S's studied are not subject to pathologically determined changes towards lesser effectiveness, or subject to degredational changes such as ageing. These assumptions confine the discussion to fictional S's. All reactive changes then are initiated to improve the S's effectiveness, however unfortunate their consequences may prove to be. Modifications to be considered in this section are sensing, correction and avoidance of error in detection problems.

A. Sensing the environment

An S, the task of which is to detect occurences of a limited class of objects, may treat all other objects as members of an inclusive equivalence class of the irrevelant. By eliminating operations and channels concerned with the recognition of irrelevant types, the decoding structure can often be considerably simplified, for example by back blocking. In general, however, irrelevant samples will enter S. Therefore operations distinguishing between relevant and irrelevant types can not be eliminated, as they can when fore testing or echo testing a particular sample from the full environment.

In the detection mode S operates on many samples, most of which are only partially processed. Samples may be discarded from further processing whenever an operation attempts to direct them to a blocked channel. These channels may be regarded as discard channels rather than as being physically blocked. The high rate of internal discard suggests that the input facilities at the root of the decoding tree may become overloaded and be a bottleneck to an otherwise adequate structure. For this reason it may be advantageous to supply the structure with multiple input facilities, so that a number of samples may be simultaneously carried through the operations of nodes occupying the first few ranks of the structure. In S's where most samples will be discarded before reaching recognition, multiple sample rather than single sample processing can be advantageous.

The difference in cost between missing an example of the class to be detected, and identifying as an example objects in fact irrelevant, will often affect choice of detection strategy. If false alarms are very costly, S will benefit by a final, precise test,

such as testing of supplementary information in a terminal checking structure. Usually few samples can be expected to reach terminal nodes. It may be relatively efficient to have the testing at lower ranks rather looser than might be desirable in the general identification mode. Looseness in preterminal processing increases the likelihood that not even atypical samples of the classes to be detected will be overlooked.

A partner to this looseness in preterminal testing is direct anticipation of terminal matching. If the class of purpose-relevant types consists of a few common and many more uncommon types, detection processing may become more efficient by first conveying the sample to the common termini for attempted matching, and resorting to the more orderly testing sequences only if this direct anticipation fails.

Detection problems differ in the degree to which the S may increase its detection rate by selectivity in its exposure to the environment. This selective exposure occurs, for example, when a person looking for a particular object moves either his person, his receptor orientation, or his attention from one part of the environment to another. When a detection procedure makes use of such responses to increase the probability of its exposure to task relevant objects, it is using search responses, a special variety of trial response. The simplest search is that pattern of environmental exposure insuring that eventually the environment will be covered with as little repetition as possible. Such a search is a scan. More interesting searches take advantage of regularities existing in the distribution of the objects of interest in the environment; the term hunt would be appropriate for these.

Searching procedures may make use of different reactive modifications in S than those called for in the simple detection mode. Thus in a hunt the class of relevant objects for detection may change as the hunt progresses. No matter how limited the class of objects for terminal detection, S may do well in early stages to recognize a much broader class of potential cues, many perhaps having little similarity to the objects sought. Examples of this progressive narrowing of the class of objects to be detected occur in information retrieval. Both search and detection, unlike the other processing modes hitherto considered, seem to require temporary memory of a rather special kind.

B. The Nature of Errors

This subsection concerns error in the detection procedures of S, and those reactive modifications in S that reduce the severity of the consequences of error. Before considering these modifications, some attention will be given to characteristics of errors themselves.

First, an error of identification may have its source in either S or in its environment. Thus an error in reading a printed page may stem from a misprint in the text or from error in the recognition process. Environmental errors may be treated as part of the variability naturally occuring in the types to be recognized. "Error" correcting transformations may be introduced as reactive modifications in the S. Such modifications might well, for instance, be based on experienced probability distributions which reflect the relationship between the environmental processes which produce the sample and the kinds of error to which these processes are vulnerable. For example, a typist may frequently hit a key adjacent to the correct one, or transpose neighboring letters. Such common kinds of error could be incorporated in transformations applied to the sample before or after an attempted terminal match.

At a deeper level the problem of distinguishing between environmental error and S processing error goes beyond questions of type variability. If a particular object has never before been recognized by S as an example of a type, S may not have formed an appropriate transformation to reduce samples of "environmental error" to a standard form appropriate for terminal matching. Faced with what seems to be an unusual object, S must decide whether the apparent peculiarity of the object stems from a failure on the part of S to recognize the familiar, or from a genuinely different quality of the object for which it has as yet no name. Most environments can be described in an infinity of ways, and the fact that there is something unusual in the grouping of environmental attributes constituting a sample is, in itself, no indication that the S need make provision for future recognition of such samples; they may never recur, or the may be of so little importance to the purposes of the S that it would do well never to attend

to such samples should they recur. Seen this way, the internal evidence of peculiarity in an S, can mean: internal error, environmental variability of a familiar type (or "error"), or novelty, important or not. There may be no way for the S to be able to determine at a given time which of these alternatives fits the sample, the processing of which has had unexpected results.

What is to be considered error may depend on the momentary purpose of the S. Some purposes may require identification so crude as to be functionally erroneous for another purpose. Thus S may not be able to establish any single error identification or correction policy of value independent of the task to which the S is momentarily committed. Costs implicit in various errors are variable and depend on purpose. Depending on the S's purpose, the occurrence of an error may not warrant correction; if it does not, sensitivity of S to the fact that error has occurred might only distract and so hinder the effectiveness of the S's performance. Thus a person in early stages of learning a foreign language may have trouble deciding whether to look up a misinterpreted word or to go ahead in the expectation that context will supply the missing significance. At a more advanced stage of learning a similar error might be symptomatic of an important gap in his understanding of the language and call for a careful diagnostic retrace to determine the source of misunderstanding. The leisure to correct specific errors may be available only to S's sufficiently developed so as to make comparatively few such errors.

Errors differ also in how conspicuous they may be to the error-making S. Often the possibility of correcting the error will depend on the precision with which the particular faulty operation can be localized, which in turn may depend on the speed with which the S becomes aware that an error has been made. When errors are likely not to be sufficiently conspicuous to the S, a teacher may be helpful in pointing them out. The effectiveness of teaching machines depends in no small degree on the immediacy with which the fact of error is communicated to the student. Especially helpful are teaching processes which communicate not only the fact of error but which identify the correct response, also. When this occurs in an S utilizing a

decoding tree the correct and the incorrect responses can be backed up until they meet at a node where the error probably originated. Ways in which adaptive S's can learn without direct environmental teaching aid are of great interest in connection with S's capable of discoveries predictable by no teacher.

C. Detection and Correction of Error

Detection of error may be signalled by a failure of matching in a terminal checking S, or by failure of concurrent checking processes, such as admittance testing. When the correct response is not known at the same time that error is detected, location of error may require retracing processes. In retracing, S tests various nodes of possible error in a more or less orderly way. The sample is transferred to different nodes from each suspicious node in turn until a correct terminus is reached. If there is no information pointing to one node as more likely to be at fault than any other, a retrace strategy which tries out nodes from the highest ranking on down is suggested. This is the preferable direction because the higher ranking nodes are, nearer to termini, which allows the elimination of a greater number of error possibilities for any given number of retrace steps. Also, should secondary errors arise in the retrace process, there are for high ranking nodes fewer dependent branch points, each perhaps calling for extensive retraces ot its own.

Retrace difficulties will be discussed in connection with the optimal ordering of tests. (See Section VII) It may be immediately seen that retrace difficulties are greater, the lower the rank of the node at which the error occurred. Accordingly, low ranking nodes must optimally be of very high accuracy. Other ways of avoiding the pyramiding difficulties of retraces back to low ranking nodes, are techniques to determine nodes of likely error without calling for a complete retrace. Attention will now be directed to these.

D. Error Avoidance

Error, whether occuring near the beginning or end of a path, will lead to terminal error. Where the error occurs, however, will have important influence on selection of error avoidance and error correction technique. Each sample may be tagged with a path uncertainty-vector, the elements of which are estimates of the probable error at each preceding nodal decision. A path uncertainty vector can be used to determine optimal retrace strategies. The path uncertainty vector may simply acquire an estimated uncertainty element as each node is passed, the values of such elements remaining fixed subsequently. Or in more interesting structures the information in the vector may be updated by input to it from on-going processes throughout the structure. This updating can simplify error correction by anticipating how error might come about, before error is detected or indeed before a terminus is reached by the critical recognition procedure. Some auxiliary checking techniques have already been mentioned: fore-processing, echo-processing and after-processing.

After-processing, as a checking procedure, may allow determination of error at a node after the sample has been transferred to a higher ranking node. Low transfer thresholds can contribute sufficiently to the overall processing speed of a structure that time lost in error correction may be offset by this speedup of testing. However when low threshold (i.e. loose) testing procedures are used, after-processing and other error locating techniques become relatively more important.

Another error-avoidance technique is the use of counter-processing. This is essentially a distributed sample structure, in which at certain nodes of possible error the sample is transferred not only to the test-selected node but to other nodes as well, despite the test. Counter-processing could be said to function as a devil's advocate in trying possibilities, individually unlikely but dangerous if correct. The term is intended to designate something more selective than widespread sample distribution with rejection in most branches. By dangerous alternatives are meant those which, if correct, lead to responses incompatible in outcome values with those following an erroneous decision. Costs of error are normally a function of the erroneous node's rank. As has been shown, retrace will generally be easier, the higher the rank of the node of error. Because of this, counter-processing, and after-processing will usually contribute more at low than at high ranking nodes.

Another error avoiding method uses multiple structures for detection. Thus if S has two detection trees, operating simultaneously on the input samples, they could serve to check each other. The set of types still possible at any interim processing stage, will be called the consequence set reached by that structure at that stage. As processing proceeds, the consequence set generally will become smaller and smaller until, at recognition, only one type is included. If S is processing the sample simultaneously in more than one structure, the logical product of the consequence sets of each of the structure can be formed. When this product is reduced so that only one type is left, recognition can be regarded as completed even if no one structure has carried processing far enough to have achieved identification by itself. This identification by first unique consequence set is a form of parallel operations available for multiple structures that can allow quick identification. Post-unique continuation of consequence set testing offers an error checking procedure. If error has occurred in any of the structures, the consequence set may become disconnected, (i.e., the logical product vanishes); no type remains that could be the outcome of all procedures.

Checking by post-unique testing of the consequence set of multiple structures will be sensitive to the extent that each different structure tests different attributes, as uncorrelated as possible. If the decisions made at corresponding nodes are correlated, the structures become more and more replicas of each other. Here the consequence set does not converge to uniqueness appreciably before the termini of the structures are reached. Highly correlated structures offer little possibility for checking other than the obvious increase in reliability to be obtained by replication of processing mechanisms.

Location of the source of error may also be accomplished in local link-forming structures. These structures allow a quite different way of path determination in which, instead of the sample being passed from node to node, from root to terminus, the sample, either as a whole or disassembled, is input to all or many nodes

simultaneously. Links are then established between nodes according to the results of testing within the nodes in question, but not necessarily in order of increasing rank. These structures are of particular interest in non-tree connectivities[5], but may be useful in trees also.

Local link forming processes can link nodes that could not have been approached in a progressive tracing structure if the lowest ranking node of the chain had been missed. Thus, there may be formed by these processes, *unrooted* node linkages, or *chains*, which may reach a terminal node and have higher overall probability of correctness than a rooted sequence produced by progressive tracing. Figure I shows how such a chain can be interpreted as "pointing" to the node at which the sequential tracing process most probably was in error.

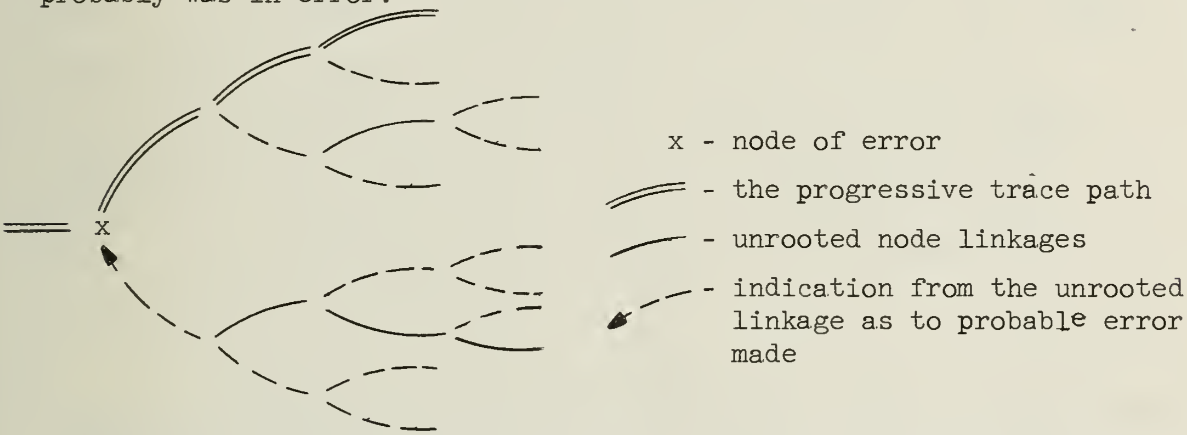

x
x - node of error
- the progressive trace path
- unrooted node linkages
- indication from the unrooted linkage as to probable error made

Figure 1: Unrooted Linkage Pointing to a Node of Probable Error

E. Undermining and Urgency

Before shifting attention from problems related to error, it may be of interest to think about reactive modification in a detection tree, when there is evidence that an error may have been committed. In some cases it will be best for the S to abandon the present procedure at once and to institute appropriate error retrace or other correction processes. The nearer the on-going process is to a relative closure,

[5] The "flash thru" feature of the Pattern Articulation Unit, originally suggested by the author on psychological grounds, is an example of a local link forming process. See Digital Computer Laboratory Reports No. 122, 125 by B. H. McCormick.

such that it can be resumed without loss if it later appears that no error occurred, the more desirable it will be for the process to be continued, at least until the point of relative closure is reached. The best point of relative closure is, of course, either a point of admittance test completion, or a terminus where the issue of correctness can be quickly resolved. When there is no proximity to a point of relative closure, continuation of the procedure will depend on: cost of the process so far, value realizable if it should prove correct, cost should it be incorrect, cost of resuming it after discontinuation should resumption be necessary, and, of course, probability that the process is in fact in error.

The greater the probability that a procedure is in error, despite apparent success in recent local tests, the more it can be said to be undermined. Undermining may be of several kinds: error undermining is caused by the possibility of downstream error,[6] although the purpose of the procedure remains unchanged; purpose undermining occurs when the doubt comes not from possible downstream error in achieving this purpose, but from the fact that, error or not, achievement of the purpose itself is losing its value as the S shifts to other purposes.

A procedure, however undermined, may undergo a reactive modification designed to reach relative closure before being terminated. This relative closure, if achievable, throws away rich possibilities or greater precision realizable had the procedure not been undermined. This urgency mode may, in different circumstances of undermining and in different S's have a variety of characteristics. Possible changes include: increased use of anticipation for the direct testing of high probability outcomes of relative closures; dropping more unlikely branches of sequencing nodes, so that the sequence exploration can be shortened; increased looseness in testing thresholds for the more probable alternatives; momentarily increased interruption threshold for the entire procedure, to compensate for

[6] That is, toward the root of the tree.

more likely interruption as additional undermining information becomes available; preparation of residua, to allow return to be as cheap as possible; and closure-forcing selectivity of testing, directed to the immediate elimination of as many as possible unlikely alternatives of procedure (for example, to aid anticipation). The urgency mode intentionally resembles the psychological phenomena of anxiety; both exhibit a reactive narrowing of on-going processes in the face of threat or suspected undermining. The urgency mode is the first example in this report of a kind of inertia or lag in changing procedures. Study of such decisional hysteresis that is optimal in different situations might prove rewarding.

The kinds of error considered hitherto have been particular nodal faults. More general errors are systemic. Diagnostic procedure may be utilized by S to test its operating effectiveness aside from any particular problem. The possible value of using intentionally introduced error for diagnostic purposes may be mentioned. If such error is found to cause no difficulties, this may be evidence that the structure is encumbered by artifactual complexities no longer necessary for its function. Thus, important simplication may be possible. Introduced error (e.g., marginal checking) may also point to possible weaknesses of the structure which unless improved may be the source of errors at awkward future times. Further discussion of diagnosis of systemic error must await a proper treatment of the subject of developmental change.

VII. DETERMINANTS OF SHORT RANGE EFFECTIVENESS

This section considers the relative reactive and adaptive effectiveness of possible S structures. The determinants of effectiveness are potentially many, few of which can be adequately specified in abstract discussion. Principles presented here will be too qualitative to allow prediction of absolute differences in effectiveness of alternate S designs. In real S's, full realization of any one of these principles of effective design will often be at the cost of other principles, the relative importance of which will depend on the specific S and task.

A. Order of Tests

The order in which tests are arranged in detection trees may seem of little consequence. Procedures end with the same types, each the logical product of all the consequence sets produced by decisions between root and terminus. Logical products are commutative, and the total information derived is a function only of the number of types and of the relative frequency of occurence of their corresponding objects in the relevant environment of the S.

If all samples called for the same tests, nothing would be gained by fixing order of testing. There might be an advantage to selecting the order for a sample by methods of queueing theory, on the basis of the momentary load on each testing facility.

However, when the range of types to be identified is large, reducing the number of tests to be performed, generally increases efficiency. Thus, once one has determined that an animal has four hooved feet, it is no longer important to ask about the number of spines in his dorsal fin, or to ask what language he speaks.

The <u>test economy</u> principle can be stated: the more generally applicable a test, the lower the rank to which it should be assigned; or, the more decisions to which a given decision is relevant, the sooner that decision should be made. Such a test can be regarded as a shared subprocedure which, done once, makes possible the processing of other subprocedures, perhaps simultaneously (as in distributed sample structures). Or again, the completion of such a test may be viewed as a way of eliminating with one test as many incorrect alternatives as possible, where this decision, if unmade, would be a component of many subprocedures of higher rank. Because low rank tests will be used on more samples than high ranked tests, it may be desirable either that they take little time, or that multiple low rank structures be provided by the S.

There may be circumstances in actual S's that prevent full realization of the test economy principles. For example, the low ranking nodal tests may be largely confined to aspects of the input data first acquired. There may be statistical support for the assignment

of low rank on the basis of seniority. Early discriminations may be, but are not always, of relatively universal relevance, but they are not always so. Even in S's capable of change, the lower the rank of the node to be changed, the greater the number of secondary changes that may be necessitated in higher order nodes. Similarly node replication for parallel processing may increase the difficulty of change.

Another limitation on the universal applicability of test economy comes from the high cost of error retracing for low ranked nodes. This leads to the rank of error principle: that the greater the frequency of error in a test, the higher the rank of the node to which it should be assigned. Statistically this may also tend to be correlated with range of applicability and seniority, as increased experience with a test may lead to its being more reliably performed. Also, inaccurate tests will often be those which need changing, and change is easier in higher ranks.

This requirement of greater accuracy in lower rank nodes suggests also that the error avoiding techniques mentioned in the preceeding section will be better employed in the lower than in the higher ranking parts of the tree. As more nodes are interjected between root and terminus in a detection tree, the structure shows greater vulnerability to error. If the sequential decisions required are statistically independent, the probability that the whole procedure has reached the correct terminus is the product of the probabilities for correctness of each test. If many decisions are involved, the overall probability of error may be high even though individual tests are of high accuracy. This observation puts a severe limitation on the depth (mean terminal rank) of decoding trees. Again this limitation may be in conflict with the economy principle..

If a test has been found to be a source of frequent error, what can be done about it? At the expense of test economy it can be assigned to higher ranking nodes, where, as the termini are close, error when it occurs will be more rapidly detected. This will

require that the test be assigned to at least one node in every descending branch. Such duplication may, however have the advantage of allowing variants of the test to be tried in each place. Should an improvement be found in the test, it may be possible to move the test back to lower rank. Other possibilities include: assignment to a node of more than one test, each normally sufficient to make the requested decisions (a multiply determined node); making the node momentarily into a sequencing node until information useful for improving the test can be acquired; and, assigning special input facilities to the node so that the sample can be augmented by additional information.

Related to these problems of accuracy are ambiguities caused by objects which are incomplete representations of their types. If due to the nature of the objects studied, particular attributes are especially likely to be missing, then tests using these attributes are better assigned to high ranked nodes of the structure, as they will often be forced to revert to sequence selection. Retrace difficulties make costly any unnecessary assignment of sequence selection to low ranked nodes. When sequencing becomes necessary at a low rank node, S must function in a less efficient way than normal, and the reactive modifications characteristic of the urgency mode may be useful.

The use of anticipation will generally be most effective close to the terminal nodes at which it can be tested. It is also most effective when the alternatives at a node are of very unequal probabilities, so that anticipatory testing of the most likely alternatives can often permit bypassing of time-consuming test discrimination among unlikely possibilities.

These two considerations can be combined in the principle that tests with very unequally probable alernatives are better assigned to high ranked nodes. Additional values of such assignment will be realized in cases where these are also nodes which serve for sequence selection. The desirability of assigning sequencing functions to high rank nodes has been mentioned. This exploitation of the inequality of alternative probabilities allows improvement over arbitrary sequencing.

Low ranked nodes, on the other hand, will generally be more effective, the more rapidly their decision processes can eliminate branches from further consideration. In information theory terms the average information conveyed by the decisions of a node is related to the average number of alternatives eliminated, and this will be greatest when the alternatives are equally probable. This is the converse of the previous principle of assigning processes with very unequal probability distributions among alternatives to high ranking nodes.

This discussion can be simplified by introducing a method of tree description, the relative ramification profile. This is constructed by plotting for each rank the ratio of the number of nodes of that rank to the number of nodes at the next lower rank, and dividing this ratio by the mean information conveyed by the decisions transferring messages between these two ranks. The two principles, relating the probability distribution among alternatives at a node to the effective rank assignment of the node, may be combined in this principle: the relative ramification profile of a detection tree should be a monotorically increasing function of the rank. High relative ramification is also associated with difficult error retrace, expecially costly at the lower ranks of the structure.

Trees in which all termini have the same rank will generally be less effective than trees which have branches of various lengths. Thus, if there are particular types that are of unusual importance to the S, or which will provide particularly valuable context information for recognition of other objects, it will be effective for the S to achieve recognition of these salient types in less time than required for other types. This can be achieved variously, but the recognition of these salient types by procedures with fewer nodes can provide not only generally higher speed, but also offer greater freedom from error because of the lesser rank depth. This kind of efficiency from saliency shortening may, unfortunately, be achievable for only a limited set of purposes, as the saliency of a type will often closely depend on purpose.

B. Segmentation by Function and by Purpose

Difficulties can arise in attempts to design one structure to be effective for more than one purpose. Ways of improving the effectiveness of structures will point to different structures for different purposes,

such as saliency shortening. When difficulty of multiple optimization is too great, S may develop partially independent structures for certain important purposes. Such an S is purpose-segmented, as contrasted to a system which is functionally segmented, as into encoding and decoding structures, both of which will be involved in achieving any purpose.

Functional segmentation can sometimes reduce the necessity of purpose-segmentation. As an illustration, let it be assumed that purpose selection has been assigned to the encoding segment (i.e. response preparation and selection structure). Grouping of termini into purpose-equivalent sets, reactive modifications of thresholds for detection, etc. will be induced from the encoding structure. This arrangement has the advantage of keeping the structure of the decoding segment free from transient changes induced by long range influences of purpose. This relative independence of the decoding segment may free its interpretation of environmental realities from excessive distortion in the service of particular purposes, an historic source of individual and social pathology. This segmentation will be most effective if the changes induced in the decoding tree diminish as the rank of the affected nodes decreases.

The above model allows some flexibility in the decoding structure to accomodate its processes to purpose. If this is not sufficient, further purpose-segmentation may still be necessary. For structures where change of purpose is frequent, reactive changes induced by interruption of purpose may so quickly change the structure that there is little opportunity to seek the shelter of relative closure or to form residua. Residua, by the way, are a particular kind of transient purpose-segments.

Trees are particularly vulnerable to demands for purpose-segmentation. If the definition of error is too closely tied to purposes, it may be impossible to confine the error-prone tests to the highest ranking nodes of the tree. This in itself may entail such difficulties in retracing that use of a single tree becomes inefficient. The utility of local link-forming structures in locating the node of

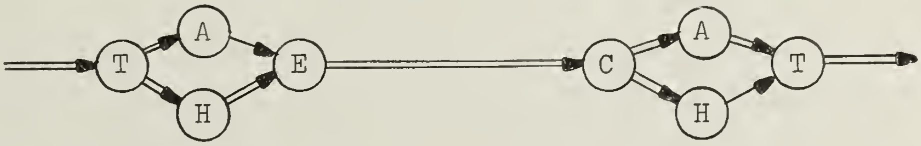
error in a progressive trace procedure has been mentioned. This technique will be of particular value in trees whose nodes can not be ranked according to accuracy. Such link-forming structures have other advantages, particularly when not tree-connected.

C. Link-Forming

Earlier mention was made of inducing reactive modifications in trees by using context to aid recognition. Most natural of these reactive modifications is the use of back blocking to simplify the decoding trees as a function of previous recognitions. The problem of how the termini to be back blocked could be selected without an auxiliary structure, perhaps as complicated as the decoding tree itself, was not met. Certain features of local link-forming structures, and in particular functionally segmented ones, that seem naturally adapted to this kind of recognition problem, will now be discussed.

Consider structures the individual nodes of which respond to possible attributes of the input samples. Part of all of the sample information is input to each node, and the node becomes activated to a degree corresponding to the likelihood, tested in the node, that the nodal attribute may be present in the sample. This activation is the initial operation of the node. Subsequent operations consist of forming links with other nodes and of multi-node operations to test the validity of particular chains. For a crude example consider a structure, the nodes of which correspond to the letters occuring in printed text. Let there be a complete set of letter nodes available for each possible letter position in the word. In Diagram II, a simple example from Selfridge,[7] the use of context in determining the recognition of an ambiguous letter is presented.

"Ⱥ" Ambiguous, i.e. A or H



[7] Selfridge, O. Pattern Recognition and Modern Computers, Proc. WJCC, 1955, p. 91-93.

The transition probabilities of English are such that "A" in the second place is seldom followed by "E", and that "H" in the second place is seldom followed by "T". Thus in this example there would be, in each case, only one path, formed from high transition probability links, leading from the first to the last position. The existence of this path might be tested by sending a flash-through pulse between the terminal positions, i.e. "T" to "T". Should there be a number of paths, they could be individually checked at the terminus against the words in a "dictionary". This matching procedure is more feasible than complete parallel matching: the number of alternatives to be tested is fewer than the number of possible permutations prior to processing in the link-forming structure. Successive increments in the threshold for link-forming as a function of transition probabilities allows testing of successively narrowed consequence sets for uniqueness, and also post-unique checking.

At the same time another local link-forming structure could be operating in which the nodes represent words, rather than letters. At this level, representation of each word possible at each sentence position would probably be uneconomical, and special constraints or tagging might be used to determine the arrangement of links. Activation of the nodes at this level would follow word recognition on the letter linking level. In addition if gaps developed in the linkages of either level, the nature of the missing information could be "pointed" to (as indicated in Figure 1) for location of error.

In this model pointing could take the form of activating certain nodes, not sufficiently activated by environmental input, to bridge the gap. It may be noted that this kind of interaction between local link-forming structures can produce fabricative transformations in which detail absent from the input is supplied from stored, past experience. This use of internal as well as external context permits very sensitive recognition on the basis of limited input, but is error prone. An example of this is what James[8] called the "proof reader's illusion". Perception of a misreading can be quite as clear as perception of correct reading, but wrong. Recently the writer saw

[8] James, W. Principles of Psychology, 2 Vol. Dover edition.

a sign on an automobile as "..... DANGER", possibly because such announcements often enough threaten some misfortune. "The sign said" .....DANCER", but the rereading produced no clearer image than the first. It is as if the first chain won the race for through connection and, once established, the details of the image were as clearly supplied from storage as from the environment.

In link forming processes, nodes belonging together become linked. Hademard[9] reports Poincare's image of thoughts being formed by the clinging together of hooked atoms mobilized in a "dance". Aside from the mobility of the nodes, this sounds much like a link-forming structure. The formal properties of the representation can be equally well satisfied by moving links between fixed entities or fixed links between moving entities.

An example of input information that may evoke incompatible purposes arises when two different objects are superimposed, for example, two different voices heard simultaneously. The purpose of detecting one may intricately interfere with the purpose of detecting the other. If they were separable prior to detection reactive modification of the structure to ignore attributes of one, or the other, would suffice. However not until detection has been accomplished will it be possible to determine which attributes were characteristic of one, which of the other. There seems to be no reasonable reactive modification of a detection tree that would make such recognition possible. Cherry[10] has shown, however, that even if two messages are read by the same voice, superimposed on tape, and input to the same ear, the human subject can reconstitute the separate messages without scrambling. This is a rather extreme example of the singular ability of human beings to make use of context in recognition. Here even purpose-segmentation would in no obvious way contribute to recognition.

[9] Hademard, J. The Psychology of Invention in the Mathematical Field, Princeton, University Press, 1949.

[10] Cherry, C. On Human Communication, Science Editions Inc, New York, 1961.

Separation of superimposed samples is possible in link-forming structures to the extent that a chain formed for one sample, fitting together its attributes, does not interfer with the simultaneous formation of a chain for the other sample. The apportionment of attributes to the proper sample, a difficult problem for a successive tracing structure, can be accomplished by attribute node linking to whichever growing tree would be better completed with that attribute in a gap. No magic vitalism is implied. This separation of superimposed samples requires something approaching essentially parallel processing; that is, the cost of alternating between subprocedures is very high. In a local link-forming structure, order depends on the success of nodes in achieving activation. This freer order can utilize particular opportunities offered in the nature of the specific input, opportunities perhaps never before anticipated.

In a tree structure, by contrast, the order of decisions is fixed. It can be well chosen according to average qualities of the inputs expected. Thus, it is conceivable that for environments which are of limited and well regulated variety, with unvarying quality of input, there may be trees which will perform more effectively than any link-forming structure. However, the difficulty with the progressive tracing structure lies in the fact that information necessary for allocation of an attribute can be obtained only after further processing which presupposes allocation. The amount of trial and error required by a progressive tracing structure is high, because decisions to be made must be accomplished in a fixed order, and progress to higher ranks occurs one step at a time.

Therefore, it is maintained here great variety in environment or in purpose can be encompassed effectively in progressive tracing structures only by costly purpose-segmentation, and can be more readily managed by link-forming structures. Fore- and back-blocking and echo testing are methods suggested for improving the effectiveness of progressive fore-testing and echo-tracing structures. These procedures attempt to alleviate difficulties inherent in fixed testing sequence by preliminary tests on sub-samples, which in turn can simplify the later critical progressive tracing process. They

effect tree simplification by local link elimination. These devices, as well as the use of multiple trees processing the same sample in parallel, in part bridge the gap between local link-forming structures and progressive tracing trees. The gap can be bridged from the other direction also, as local link forming structures can utilize tree structures when the information to be processed can be reduced by natural or artificial constraints to the simplicity required for the more sequential processing.

It was suggested earlier that efficient sequencing of a previously parallel process depends on the existence of sufficient points of relative closure for subprocedure alternation to occur without excessive costs for resumption of temporarily discontinued processes. Relative closure allows for interruption of current activity by error retracing processes. Without such closure errors would be shifty targets that would not stay put when found. Language function may use a relatively sequential, tree-like structure, richly supplied with relative closure. Important problems of a more parallel S might still be solved by such highly organized, if limited, structures as this model of language.

It is easy to picture the growth of a tree structure by the sprouting of new limbs and twigs as finer and finer discriminations become useful. The development of linking systems suggests no such facile analogies. Analysis of how growth may come about can be profoundly interesting. Understanding of developmental possibilities requires further specification of at least two kinds: i) specification of ways in which links are formed and chains tested for validity; (ii) specification of ways to coordinate sub-structures such as interacting letter-level and word-level linking planes, or such as progressive tracing fields and linking fields.

In the only technique of link formation so far suggested, the probability of linkage reflects some accumulated estimate of transition probabilities. This possibility follows the well-worn tradition of associationist models of brain function. It would be surprising if such a plausible principle were to be shown to have nothing

whatever to do with brain mechanisms. On the other hand, it would be very surprising if other determinants of linking could not be conceived, found useful in design of various S's, and perhaps identified in existing S's. Determinants of linking may reflect non-probabilistic conventions. For example, McCormick[11] has suggested that the syntactic rules of languages may operate through link-formation. Link-forming in one part of a structure is conceived here as potentially producing a vast number of different linkages. Which will occur and endure is in large part determined by information input from the environment and from other parts of the S.

A simple example of a model in which influences external to S can affect the selection of links is one in which there are distinct start nodes and distinct goal nodes, the task being to grow linked intervening nodes. Let the start nodes represent states that can be readily reached by the S in its present environmental and internal circumstances, and the goal nodes represent states in which purposes valuable to the S have been achieved. Linkage between a start node and a goal node through intervening nodes indicates to the S that the goal can probably be reached in the current dircumstance, as each is linked only with other nodes ultimately reachable from the state represented by the start node.

The efficacy of this selection is based in part on the fact that input to the nodes will activate only those relevant to the situation, and in part on the further selection of only those potentially linked states appropriate to input and purpose. In this model linkages are regarded as being directed, from a start node or from a goal node. Add the requirement that there be no cycles in the linkages between nodes, and the resulting linkages are partially ordered. There will grow, then, two trees, one rooted in the start node, the other in the goal node. When ultimately contact is established between a terminus of one tree and any node of the other, linkage has been established. According to the nature of the problem and the environmental situation it may be more

[11] Private communication

effective for linkages to grow from the start node at a slower or faster rate than those from the goal node. A first elaboration of this model allows radial growth of linkages in both directions from intermediate nodes (regarded as subgoals). Note that great savings in the number of possible linkages and decisions accrue from specification of a goal (in narrowing the forward branching of the linkages from the start node), the additional saving achieved by backward branching from the goal, and the yet additional savings by growth from subgoals. Each of these economies narrows further the range over which less guided linkings might grow.

Interacting letter and word linking structures have been mentioned. Another possibility is related to fore-processing and echo-testing techniques previously described for tree structures. Suppose that the same message is input to the nodes of two link-forming structures, which may be visualized as parallel planes. Both planes have one or more start nodes to be linked to one or more goal nodes, and these linkages are directional. Let the two planes be called the strategic and the tactic plane respectively. The strategic plane is supposed to be different from the tactic one. It should be more richly interconnected, allowing freer and quicker linkages to fan out from start nodes and fan in toward goal nodes, and offering more resistance to change through learning. Suppose that there are fewer nodes on the strategic plane, each possibly linked to many others of that plane. On the other hand each of the more numerous nodes of the tactical plane is allowed to link to fewer nodes in that plane. Finally let there be linkages possible between planes such that there will be many-few connections from the tactic plane to the strategic plane, and few-many connections in the opposite direction.

Introduction of input into S will generally be accompanied by a connection between some start node and some goal node in the strategic plane before much happens in the tactic plane. Completion of this linkage suggests that opportunities exist for reaching at least one goal from one of the available start points. Let completion of the strategic linkage activate nodes in the tactic plane corresponding to the strategic start node and goal node. In this way

the strategic plane can select a particular one of many possible goal points in the tactic plane and sensitize nodes leading to it. This selection reduces the burden of satisfying purposes not realizable by the simultaneous procedures in the tactic plane. The path connecting the start and goal nodes in the strategic plane may, through activating regions in the tactic plane, help a linkage that would have been difficult with the greater reality constraint of the tactic plane with the narrower, but more rigorous, linking horizons of its nodes.

This two plane coordinated linking structure is the simplest of many such multiple linking structure models, some of which have quite interesting potentialities. Further elaboration of these models, as well as the crucial issue of developmental change, will be deferred to subsequent publication.